

LOAD DISTRIBUTING METHOD FOR CONTENTS DISTRIBUTING SYSTEM AND SERVER COMPUTER FOR THE SAME SYSTEM

Publication number: JP2002278823 (A)

Publication date: 2002-09-27

Inventor(s): TSUKADA ERI +

Applicant(s): TOSHIBA CORP +

Classification:

- International: G06F12/00; G06F13/00; G06F15/177; G06F12/00; G06F13/00; G06F15/16; (IPC1-7); G06F12/00; G06F13/00; G06F15/177

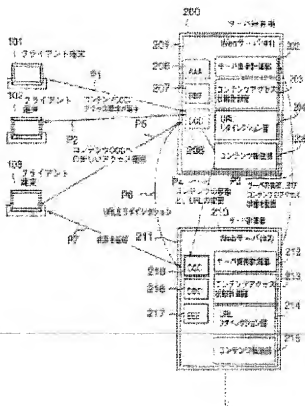
- European:

Application number: JP20010081283 20010321

Priority number(s): JP20010081283 20010321

Abstract of JP 2002278823 (A)

PROBLEM TO BE SOLVED: To distribute a load inside a system without requiring a load distributor or the like. **SOLUTION:** When the requests of access to contents CCC on a Web server 201 are concentrated (P1 and P2), loads measured by server load measuring parts 202 and 212 are compared. When the load of a Web server 211 is lighter and the frequency of access to the contents CCC is highest, a contents transfer part 205 moves the contents CCC to the Web server 211 of the light load (P3 and P4). At such a time, a URL redirection part 204 changes the URL of the contents CCC after move. In such a case, when access to the contents CCC is requested (P5), the URL redirection part 204 changes the URL in the access request to the URL of the contents CCC after move and sends the access request after that change to the Web server 211.



Data supplied from the *espacenet* database — Worldwide

(51) Int.Cl. ⁷	願明番号	F I	キーワード(参考)
G 0 6 F 12/00	5 4 6	G 0 6 F 12/00	5 4 6 A 5 B 0 4 j
13/00	5 2 0	13/00	5 2 0 D 5 B 0 8 2
15/177	6 7 4	15/177	6 7 4 B

審査請求 未請求 請求項の数7 O L (全9頁)

(21) 出願番号 特開2001-81283(P2001-81283)

(22) 出願日 平成13年3月21日 (2001.3.21)

(71) 出願人 00003078

株式会社東芝
東京都港区芝浦一丁目1番1号

(72) 発明者 塚田 絵里

東京都府中市東芝町1番地 株式会社東芝
府中事業所内

(74) 代理人 100058479

弁理士 鈴江 武彦 (外6名)

Pターム(参考) 5B045 GC04

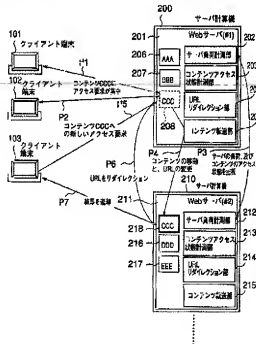
5B082 FA04 HA05 HA08

(54) 【発明の名称】 コンテンツ配信システムにおける負荷分散方法及び同システムにおけるサーバ計算機

(57) 【要約】

【課題】 負荷分散装置等を必要とすることなくシステム内での負荷分散を図ることができるようにする。

【解決手段】 Webサーバ201上のコンテンツCCCへのアクセス要求が集中した場合 (P1、P2)、サーバ負荷計測部202、212で計測される負荷を比較する。Webサーバ211の負荷の方が軽く、且つコンテンツCCCのアクセス頻度が最も高い場合、コンテンツ転送部205はコンテンツCCCを負荷の軽いWebサーバ211に移動する (P3、P4)。このときURLリダイレクション部204は、移動後のコンテンツCCCのURLを変更する。ここで、コンテンツCCCへのアクセスが要求されると (P5)、URLリダイレクション部204はアクセス要求中のURLを移動後のコンテンツCCCのURLに変更し、その変更後のアクセス要求をWebサーバ211に送出する。



【特許請求の範囲】

【請求項1】 複数のクライアント端末と、当該クライアント端末からのアクセス要求に応じて要求されたコンテンツを配信する複数のサーバ計算機とが、ネットワークにより接続されたコンテンツ配信システムにおける負荷分散方法であって、

前記サーバ計算機が当該サーバ計算機上でのコンテンツアクセス状態を監視して、アクセスが集中しているコンテンツを検出するステップと、

前記アクセスが集中しているコンテンツが検出された場合、前記サーバ計算機が当該サーバ計算機自身と隣接サーバ計算機との負荷状態を比較するステップと、

前記隣接サーバ計算機の方が負荷が低い場合、前記サーバ計算機上の前記検出されたコンテンツを当該サーバ計算機から前記隣接サーバ計算機に転送するステップと、

前記検出されたコンテンツが前記隣接サーバ計算機に転送された際に当該コンテンツのロケーション情報を変更するステップと、

前記転送元のサーバ計算機が、前記隣接サーバ計算機に転送されたコンテンツを指定するロケーション情報を含むアクセス要求を前記クライアントから受け取った場合、当該アクセス要求中のロケーション情報を前記変更後のロケーション情報に変更し、当該ロケーション情報変更後のアクセス要求を前記隣接サーバ計算機に転送するステップとを具備することを特徴とするコンテンツ配信システムにおける負荷分散方法。

【請求項2】 複数のクライアント端末と、当該クライアント端末からのアクセス要求に応じて要求されたコンテンツを配信する複数のサーバ計算機とが、ネットワークにより接続されたコンテンツ配信システムにおける負荷分散方法であって、

前記サーバ計算機が当該サーバ計算機上でのコンテンツアクセス状態を監視して、アクセスが集中しているコンテンツを検出するステップと、

前記アクセスが集中しているコンテンツが検出された場合、前記サーバ計算機が当該サーバ計算機自身と隣接サーバ計算機との負荷状態を比較するステップと、

前記隣接サーバ計算機の方が負荷が低い場合、前記サーバ計算機上の前記検出されたコンテンツを除くコンテンツを当該サーバ計算機から前記隣接サーバ計算機に転送するステップと、

前記検出されたコンテンツを除くコンテンツが前記隣接サーバ計算機に転送された際に当該コンテンツの転送元のサーバ計算機が当該コンテンツのロケーション情報を変更するステップと、

前記転送元のサーバ計算機が、前記隣接サーバ計算機に転送されたコンテンツを指定するロケーション情報を含むアクセス要求を前記クライアントから受け取った場合、当該アクセス要求中のロケーション情報を前記変更

後のロケーション情報に変更し、当該ロケーション情報変更後のアクセス要求を前記隣接サーバ計算機に転送するステップと

を具備することを特徴とするコンテンツ配信システムにおける負荷分散方法。

【請求項3】 複数のクライアント端末と、当該クライアント端末からのアクセス要求に応じて要求されたコンテンツを配信する複数のサーバ計算機と、コンテンツが予め登録されたデータベースとが、ネットワークにより接続されたコンテンツ配信システムにおける負荷分散方法であって、

前記サーバ計算機が、前記クライアント端末からコンテンツを指定するロケーション情報を含むアクセス要求を受け取った場合に、当該サーバ計算機自身と隣接サーバ計算機との負荷状態を比較し、負荷が低い方のサーバ計算機を選択するステップと、

前記クライアント端末からのアクセス要求を受け取った前記サーバ計算機が、当該アクセス要求で指定されたロケーション情報により前記データベースから対応するコンテンツを検索するステップと、

前記コンテンツを検索した前記サーバ計算機が、当該コンテンツを前記選択したサーバ計算機上に展開するステップと、

前記コンテンツを展開した前記サーバ計算機が、展開前のコンテンツのロケーション情報と展開後のコンテンツのロケーション情報とを対応付けるステップと、

前記展開されたコンテンツを指定するアクセス要求を受け取った前記サーバ計算機が、当該アクセス要求で指定されたロケーション情報を前記展開後のコンテンツのロケーション情報に変更して当該展開後のコンテンツをアクセスさせるステップとを具備することを特徴とするコンテンツ配信システムにおける負荷分散方法。

【請求項4】 複数のクライアント端末と、当該クライアント端末からのアクセス要求に応じて要求されたコンテンツを配信する複数のサーバ計算機とが、ネットワークにより接続されたコンテンツ配信システムにおけるサーバ計算機であって、

当該サーバ計算機自身と隣接サーバ計算機との負荷状態を監視・計測するサーバ負荷計測手段と、

当該サーバ計算機上でのコンテンツアクセス状態を監視して、アクセスが集中しているコンテンツを検出するコンテンツアクセス状態計測手段と、

前記コンテンツアクセス状態計測手段によりアクセスが集中しているコンテンツが検出され、且つ前記サーバ負荷計測手段による当該サーバ計算機自身と隣接サーバ計算機との負荷状態の監視・計測の結果、前記隣接サーバ計算機の方が負荷が低い場合、前記検出されたコンテンツを当該サーバ計算機から前記隣接サーバ計算機に転送するコンテンツ転送手段と、

前記コンテンツ転送手段により前記隣接サーバ計算機に

転送されたコンテンツのロケーション情報を変更すると共に、前記隣接サーバ計算機に転送されたコンテンツを指定するロケーション情報を含むアクセス要求を前記クライアントから受け取った場合、当該アクセス要求中のロケーション情報を前記変更後のロケーション情報に変更し、当該ロケーション情報変更後のアクセス要求を前記隣接サーバ計算機に転送するロケーション情報リダイレクション手段とを具備することを特徴とするコンテンツ配信システムにおけるサーバ計算機。

【請求項5】 複数のクライアント端末と、当該クライアント端末からのアクセス要求に応じて要求されたコンテンツを配信する複数のサーバ計算機とが、ネットワークにより接続されたコンテンツ配信システムにおけるサーバ計算機であって、

当該サーバ計算機自身と隣接サーバ計算機との負荷状態を監視・計測するサーバ負荷計測手段と、当該サーバ計算機上でコンテンツアクセス状態を監視して、アクセスが集中しているコンテンツを検出するコンテンツアクセス状態計測手段と、

前記コンテンツアクセス状態計測手段によりアクセスが集中しているコンテンツが検出され、且つ前記サーバ負荷計測手段による当該サーバ計算機自身と隣接サーバ計算機との負荷状態の監視・計測の結果、前記隣接サーバ計算機の方が負荷が低い場合、前記検出されたコンテンツを除く当該サーバ計算機上のコンテンツを当該サーバ計算機から前記隣接サーバ計算機に転送するコンテンツ転送手段と、

前記コンテンツ転送手段により前記隣接サーバ計算機に転送されたコンテンツのロケーション情報を変更すると共に、前記隣接サーバ計算機に転送されたコンテンツを指定するロケーション情報を含むアクセス要求を前記クライアントから受け取った場合、当該アクセス要求中のロケーション情報を前記変更後のロケーション情報に変更し、当該ロケーション情報変更後のアクセス要求を前記隣接サーバ計算機に転送するロケーション情報リダイレクション手段とを具備することを特徴とするコンテンツ配信システムにおけるサーバ計算機。

【請求項6】 複数のクライアント端末と、当該クライアント端末からのアクセス要求に応じて要求されたコンテンツを配信する複数のサーバ計算機と、コンテンツが予め登録されたデータベースとが、ネットワークにより接続されたコンテンツ配信システムにおけるサーバ計算機であって、

前記クライアント端末からコンテンツを指定するロケーション情報を含むアクセス要求を受け取った場合に、当該サーバ計算機自身と隣接サーバ計算機との負荷状態を比較し、負荷が低い方のサーバ計算機を選択するサーバ負荷計測手段と、

前記受け取ったアクセス要求で指定されたロケーション情報により前記データベースから対応するコンテンツを

検索して、当該コンテンツを前記サーバ負荷計測手段により選択されたサーバ計算機上に展開するコンテンツ展開手段と、

前記コンテンツ展開手段による展開前のコンテンツのロケーション情報と展開後のコンテンツのロケーション情報とを対応付け、前記アクセス要求で指定されたロケーション情報を前記展開後のコンテンツのロケーション情報に変更して当該展開後のコンテンツをアクセスさせるロケーション情報リダイレクション手段とを具備することを特徴とするコンテンツ配信システムにおけるサーバ計算機。

【請求項7】 複数のクライアント端末と、当該クライアント端末からのアクセス要求に応じて要求されたコンテンツを配信する複数のサーバ計算機とが、ネットワークにより接続されたコンテンツ配信システムにおいて前記サーバ計算機により実行されるプログラムであって、前記サーバ計算機に、

当該サーバ計算機上でコンテンツアクセス状態を監視して、アクセスが集中しているコンテンツを検出するステップと、

前記アクセスが集中しているコンテンツが検出された場合、当該サーバ計算機自身と隣接サーバ計算機との負荷状態を比較するステップと、

前記隣接サーバ計算機の方が負荷が低い場合、前記サーバ計算機上の前記検出されたコンテンツまたは当該コンテンツを除くコンテンツを当該サーバ計算機から前記隣接サーバ計算機に転送するステップと、

前記検出されたコンテンツまたは当該コンテンツを除くコンテンツが前記隣接サーバ計算機に転送された際に当該コンテンツのロケーション情報を変更するステップと、

前記隣接サーバ計算機に転送されたコンテンツを指定するロケーション情報を含むアクセス要求を前記クライアントから受け取った場合、当該アクセス要求中のロケーション情報を前記変更後のロケーション情報に変更し、当該ロケーション情報変更後のアクセス要求を前記隣接サーバ計算機に転送するステップとを実行させるプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、クライアントからネットワーク経由で要求されたコンテンツを要求元に配信するコンテンツ配信システムにおける負荷分散方法及び同システムにおけるサーバ計算機に関する。

【0002】

【従来の技術】 クライアントからネットワーク経由で要求されたコンテンツを要求元に配信するコンテンツ配信システムとして、Webシステムが知られている。このWebシステムの負荷分散方法としては、主として、以下の2通りがある。

【0003】第1の負荷分散方法は、複数のWebサーバ(Webサーバ計算機)上に、同一コンテンツを配置したコピーサイトを作っており、負荷分散装置やスイッチ機器といった、サーバの負荷計測機能及びリクエストの振り分け機能を持つ専用の装置により、サーバのアクセス負荷を計測し、負荷の低いサーバのコピーサイトに、クライアント(クライアント端末)からのリクエストを振り分ける方法である。

【0004】第2の負荷分散方法は、負荷分散装置やスイッチ機器といった、サーバの負荷計測機能及びリクエストの振り分け機能を持つ装置と、コンテンツのレプリケーションを動的に作成する手段とを組み合わせて、サーバの負荷を計測し、負荷の低いサーバ上の動的に作成したコピーサイトに、クライアントからのリクエストを振り分ける方法である。

【0005】いづれの方法も、Webシステム(コンテンツ)サーバの負荷を分散させることができるため、クライアントから負荷の高いWebサーバ(コンテンツサーバ)へのリクエストが発生してもクライアントへのサービスが遅れるのを防止することができる。

【0006】

【発明が解決しようとする課題】しかしながら、上記した従来技術においては、システム構成として、クライアントとサーバとの間に、クライアントからのリクエストを受け取ってサーバに振り分けを行くための専用の負荷分散装置等が必要となるという問題があった。また、予め複数のサーバに同一コンテンツを作成したり、コピーサイトとマスターサイトとの更新の同期をとる必要があるなど、コンテンツ管理が煩雑になる等の問題もあった。

【0007】本発明は上記事情を考慮してなされたものでその目的は、負荷分散装置等を必要とすることなくシステム内での負荷分散を図ることができるコンテンツ配信システムにおける負荷分散方法及び同システムにおけるサーバ計算機を提供することにある。

【0008】

【課題を解決するための手段】本発明は、複数のクライアント端末と、当該クライアント端末からのアクセス要求に応じて要求されたコンテンツを配信する複数のサーバ計算機と、ネットワークにより接続されたコンテンツ配信システムにおける負荷分散方法であって、上記サーバ計算機が当該サーバ計算機上でのコンテンツアクセス状態を監視して、アクセスが集中しているコンテンツを検出するステップと、上記アクセスが集中しているコンテンツが検出された場合、上記サーバ計算機が当該サーバ計算機自身と隣接サーバ計算機との負荷状態を比較するステップと、上記隣接サーバ計算機の方が負荷が低い場合、上記サーバ計算機上の上記検出されたコンテンツを当該サーバ計算機から上記隣接サーバ計算機に転送するステップと、上記検出されたコンテンツが上記隣接

サーバ計算機に転送された際に当該コンテンツの転送元のサーバ計算機が当該コンテンツのロケーション情報を変更するステップと、上記転送元のサーバ計算機が、上記隣接サーバ計算機に転送されたコンテンツを指定するロケーション情報を含むアクセス要求をクライアントから受け取った場合、当該アクセス要求中のロケーション情報を上記変更後のロケーション情報に変更し、当該ロケーション情報変更後のアクセス要求を上記隣接サーバ計算機に転送するステップとを備えたことを特徴とする。

【0009】このような構成においては、あるサーバ計算機上のコンテンツにアクセスが集中した場合に、そのサーバ計算機と隣接サーバ計算機の負荷が比較され、隣接サーバ計算機の負荷の方が軽い場合、その負荷の軽い隣接サーバ計算機に当該コンテンツが転送されると、その当該コンテンツのロケーション情報を有するサーバ計算機のコンテンツへの新たなアクセス要求がロケーション情報を変えて隣接サーバ計算機に転送される。これにより、アクセスの集中したコンテンツを有するサーバ計算機の負荷を下げ、システム全体のサーバ計算機の負荷を均等にし、システム全体の応答性を高めることができる。しかも、負荷分散装置等を使用することなく、サーバ計算機の負荷を均質化できる。

【0010】ここで、アクセスの集中したコンテンツを転送する形態として、当該コンテンツを移動する構成とするならば、コンテンツの一元管理が簡単に行える。

【0011】また、アクセスの集中したコンテンツを転送するのではなく、アクセスの集中したコンテンツは残し、それ以外のコンテンツを負荷の軽いサーバ計算機に転送する構成とするならば、接続中のクライアントに対し、ロケーション情報の変更等の影響が少なくて済む。

【0012】また本発明は、各サーバ計算機に予めコンテンツを置く代わりに、コンテンツが予め登録されたネットワーク経由でアクセス可能なデータベースを用意し、各サーバ計算機で実行される負荷分散方法として、上記サーバ計算機がクライアント端末からコンテンツを指定するロケーション情報を含むアクセス要求を受け取った場合に、当該サーバ計算機自身と隣接サーバ計算機との負荷状態を比較し、負荷が低い方のサーバ計算機を選択するステップと、クライアント端末からのアクセス要求を受け取ったサーバ計算機が、当該アクセス要求で指定されたロケーション情報により上記データベースから対応するコンテンツを検索するステップと、上記コンテンツを検索したサーバ計算機が、当該コンテンツを上記選択したサーバ計算機上に展開するステップと、上記コンテンツを展開したサーバ計算機が、展開後のコンテンツのロケーション情報と対応付けるステップと、上記展開されたコンテンツを指定するアクセス要求を受け取ったサーバ計算機が、当該アクセス要求で指定されたロケーション

情報を上記展開後のコンテンツのロケーション情報に変更して当該展開後のコンテンツをアクセスさせるステップとを備えたことを特徴とする。

【0013】このような構成においては、サーバ計算機の負荷を均等にすることによる、システム全体の応答性能を高めると共に、コンテンツの一元管理を行うことができる。また、コンテンツの管理をデータベースを用いて行うことにより、安全性も向上する。

【0014】以上の構成のコンテンツ配信システムにおける負荷分散方法に係る本発明は、当該負荷分散方法を実行するサーバ計算機に係る発明としても成立し、また当該負荷分散方法を適用し、且つサーバ計算機で実行されるプログラムの発明としても成立する。

【0015】

【発明の実施の形態】以下、本発明のコンテンツ配信システムをWebシステムに適用した実施の形態につき図面を参照して説明する。

【0016】【第1の実施形態】図1は本発明の第1の実施形態に係るWebシステムのシステム構成を示す。

【0017】図1のシステムにおいて、複数のクライアント端末、例えばクライアント端末101～103と、N台のサーバ計算機200、210、…220とは、ネットワーク300により相互接続されている。

【0018】サーバ計算機200、210、…220上には、それぞれクライアントからのHTTP (Hypertext Transfer Protocol) によるリクエスト (アクセス要求) を受け付けて結果をリクエスト元に返答するWebサーバ (＃1、＃2、…＃N) 201、211、…221が構築されている。この場合、サーバ計算機200、210、…220自体をWebサーバと呼ぶこともある。

【0019】Webサーバ (＃1、＃2) 201、211は、自身 (自サーバ) と隣接サーバとの負荷状態を監視・計測するサーバ負荷計測部202、212と、コンテンツアクセス状態を監視し、アクセス頻度が最も高いコンテンツを検出するコンテンツアクセス状態計測部203、213とを有している。ここでは、Webサーバ (＃1) 201とWebサーバ (＃2) 211とは隣接サーバであるものとする。隣接サーバとは、図1のWebサーバ201とWebサーバ211の例のように物理的に隣接している場合に限らず、例えば各Webサーバが有するシステム構成Webサーバのリスト上で隣接しているものであっても構わない。

【0020】Webサーバ (＃1、＃2) 201、211はまた、自サーバ上のコンテンツが他のサーバ上に移動された際に、そのコンテンツの所在位置を示すURL (Uniform Resource Locator) と呼ばれるロケーション情報を動的に変更するURLリダイレクション部204、214と、自サーバ上のコンテンツを指定のサーバに転送するコンテンツ転送部205、215とを有して

いる。

【0021】Webサーバ (＃1、＃2) 201、211上のサーバ負荷計測部202、212、コンテンツアクセス状態計測部203、213、URLリダイレクション部204、214及びコンテンツ転送部205、215は、サーバ計算機200、210上で所定のプログラムを実行することにより実現される。

【0022】Webサーバ (＃N) 221もまた、Webサーバ (＃1、＃2) 201及び211と同様の構成を有している。

【0023】また、Webサーバ201には、コンテンツ (AAA, BBB, CCC) 206、207、208が配置 (格納) されている。ここでは、コンテンツ (CCC) 208が、コンテンツ (AAA, BBB, CCC) 206、207、208のうちで最もアクセス頻度が高いものとする。一方、Webサーバ211には、コンテンツ (DDD, EEE) 216、217が配置 (格納) されている。

【0024】次に、図1のシステムにおける動作を図2のシーケンスチャートと参照して説明する。まず、クライアント端末101、102から、Webサーバ201上のコンテンツCCC (208) にアクセス要求が集中し、コンテンツCCC (208) へのアクセス頻度が予め定められた閾値を超えているものとする (ステップP1、P2)。

【0025】Webサーバ201、202は互いに隣接している。このためWebサーバ201、202のサーバ負荷計測部202、212により、自サーバ201、211の負荷状態を監視・計測すると共に、隣接サーバ211、201の負荷状態を問い合わせ、いずれの負荷の方が高いかを判断する (ステップP3)。ここで、サーバの負荷状態には、対応するサーバ計算機の単位時間 (一定周期) 毎のCPUの稼働時間が用いられる。上記ステップ3では、Webサーバ201、211は、自身の方が負荷が高い (重い) と判断した場合、コンテンツアクセス状態計測部203、213により、自サーバ上に配置されているコンテンツの中でアクセス頻度の最も高いコンテンツを検出する。

【0026】ここでは、Webサーバ201の方がWebサーバ211より負荷が高いと判断され、その結果、Webサーバ201上のコンテンツアクセス状態計測部203により、アクセス頻度の最も高いコンテンツとしてコンテンツ (CCC) 208が検出されたものとする。

【0027】この場合、Webサーバ201はコンテンツ転送部205により、コンテンツ (CCC) 208を負荷の低いWebサーバ211にコンテンツ (CCC) 218として転送、例えば移動する (ステップP4)。このステップ4では、Webサーバ201は更にURLリダイレクション部204により、移動後のコンテン

ツ(CCC)218のURLの変更を動的に行う。URLリダイレクション部204では、移動前のコンテンツ(CCC)208のURLと移動後のコンテンツ(CCC)218のURLとの対応表を保持する。

【0028】このような状態で、例えばクライアント端末103からWebサーバ201に対してコンテンツ(CCC)208への新しいアクセス要求が与えられたものとする(ステップP5)。

【0029】するとWebサーバ201上のURLリダイレクション部204は、移動前のコンテンツ(CCC)208と移動後のコンテンツ(CCC)218とのURL対応表に使い、アクセス要求中のURLを移動後のコンテンツ(CCC)218を指す新しいURLに変更し、当該URL変更後のアクセス要求をWebサーバ211に送出する(ステップP6)。

【0030】これによりWebサーバ211は、変更後のコンテンツ(CCC)218を結果のHTMLとしてクライアント端末103に返信する(ステップP7)。なお、Webサーバ211に接続するWebサーバ201とは別のWebサーバの方が負荷が低い場合には、コンテンツ(CCC)218は当該別のWebサーバに移動されると共にURLが変更される。そしてクライアント端末103からのコンテンツ(CCC)208へのアクセス要求は、URLを書き換えてWebサーバ211に転送され、更に当該Webサーバ211でURLが書き換えられて上記別のWebサーバに転送されることになる。

【0031】このように本実施形態においては、アクセスの集中したコンテンツを負荷の軽い他サーバに移動することにより、サーバの負荷を均質化し、システム全体としての応答性能を向上させることができる。しかも、システム構成として、負荷分散装置を使用せずに、サーバの負荷を均質化できる。更にコンテンツのコピーではなく、移動を行うため、コンテンツの一元管理が行える。ここで、コンテンツを移動する際に当該コンテンツにアクセス中のクライアントが存在する場合には、当該コンテンツのアクセスが終了してから当該コンテンツを移動する。この他、コンテンツの移動ではなくコンテンツの複写(コピー)を行うならば、コンテンツの一元管理はできないものの、当該コンテンツにアクセス中のクライアントが存在しても、当該コンテンツを速やかに複写できる。

【0032】なお、本実施形態においては、あるコンテンツへのアクセス頻度が閾値を超えたならば、負荷の軽いWebサーバに当該コンテンツを移動(複写)してURLを変更するものとして説明したが、これに限るものではない。例えば、あるコンテンツへのアクセス頻度が閾値を超えても、対応するWebサーバの負荷が予め定められた負荷(閾値)に達していないならば、当該コンテンツを移動(複写)しないようにすることも可能であ

る。これにより、必要以上にコンテンツの移動(複写)が発生して、クライアントからのアクセス要求が転送されるのを防止できる。

【0033】〔変形例〕以上に述べた実施形態では、アクセス要求が集中したコンテンツを移動(複写)するものとして説明したが、これに限るものではない。そこで、図1のシステムにおいて、アクセス要求が集中したコンテンツを残し、その他のコンテンツを負荷の軽いサーバに移動(複写)するようにした変形例について、図3のシーケンスチャートを参照して説明する。

【0034】まず、クライアント端末101、102から、Webサーバ201上のコンテンツCCC(208)にアクセス要求が集中し、コンテンツCCC(208)へのアクセス頻度が予め定められた閾値を超えているものとする(ステップQ1、Q2)。

【0035】Webサーバ201、202はサーバ負荷計測部202、212により、自サーバ201、211の負荷状態を監視・計測すると共に、隣接サーバ211、201の負荷状態を問い合わせ、いずれの負荷の方が高いかを判断する(ステップQ3)。このステップQ3では、Webサーバ201は、自身の方が負荷が高いと判断した場合、コンテンツアクセス状態計測部203により、自サーバ上に配置されているコンテンツの中でアクセス頻度の最も高いコンテンツを検出する。

【0036】ここでは、Webサーバ201の方がWebサーバ211より負荷が高いと判断され、その結果、Webサーバ201上のコンテンツアクセス状態計測部203により、アクセス頻度の最も高いコンテンツとしてコンテンツ(CCC)208が検出されたものとする。

【0037】この場合、Webサーバ201はコンテンツ転送部205により、アクセス頻度の最も高いコンテンツ(CCC)208以外のコンテンツ、即ちコンテンツ(AAA)206及びコンテンツ(BBB)207を負荷の低いWebサーバ211にコンテンツ(AAA)226及びコンテンツ(BBB)227として転送、例えば移動する(ステップQ4)。このステップQ4では、Webサーバ201は更にURLリダイレクション部204により、移動後のコンテンツ(AAA)226及びコンテンツ(BBB)227のURLの変更を動的に行う。URLリダイレクション部204では、移動前のコンテンツ(AAA)206及びコンテンツ(BBB)207のURLと移動後のコンテンツ(AAA)226及びコンテンツ(BBB)227のURLとの対応表を保持する。

【0038】以降の動作は図2のシーケンスチャートを参照して説明した前記実施形態と同様であるため、説明を省略する。

【0039】このように本変形例においては、アクセスの集中したコンテンツを残し、その他のコンテンツを負

荷の軽いサーバに移動することにより、サーバの負荷を均質化し、システム全体としての応答性能を向上させることができる。しかも、システム構成として、負荷分散装置等を使用せずに、サーバの負荷を均質化できる。更にアクセスの集中したコンテンツを残し、その他のコンテンツを移動するため、サーバに接続中のクライアントに対し、URL等の変更の影響が少なくて済む。

【0040】〔第2の実施形態〕図4は本発明の第2の実施形態に係るWebシステムのシステム構成を示す。【0041】図4のシステムにおいて、複数のクライアント端末、例えばクライアント端末401、402と、N台のサーバ計算機500、510、…520と、データベース600とは、ネットワーク700により相互接続されている。

【0042】サーバ計算機500、510、…520上には、それぞれクライアントからのHTTPによるリクエスト（アクセス要求）を受け付けて結果をリクエスト元へ返信するWebサーバ（#1、#2、…#N）501、511、…521が構築されている。

【0043】Webサーバ（#1）501は、自身（自サーバ）と隣接サーバとの負荷状態を監視・計測する（図1中のサーバ負荷計測部202と同様の）サーバ負荷計測部502と、アクセス要求で指定されたURLに従って対応するコンテンツが自サーバ上に展開された場合に、当該URLを展開されたコンテンツのURLに変更するURLリダイレクション部503と、コンテンツ展開部504とを有している。コンテンツ展開部504は、アクセス要求で指定されたURLに従って対応するコンテンツを自サーバ上に展開する。

【0044】Webサーバ（#1）501上のサーバ負荷計測部502、URLリダイレクション部503及びコンテンツ展開部504は、サーバ計算機500上で所定のプログラムを実行することにより実現される。

【0045】データベース600には、コンテンツ（AAA）601など、クライアント端末401、402からのアクセス要求中のURLで指定される種々のコンテンツ（コンテンツデータ）が登録されている。

【0046】Webサーバ（#2）511及びWebサーバ（#N）521もまた、Webサーバ（#1）501と同様の構成を有している。

【0047】次に、図4のシステムにおける動作を図5のシーケンスチャートを参照して説明する。まず、クライアント端末401からWebサーバ（#1）501に対し、コンテンツAAAを指定するURLへのアクセス要求が送出されたものとする（ステップR1）。

【0048】Webサーバ501は、クライアント端末401からのアクセス要求を受け取ると、サーバ負荷計測部502により自サーバ501の負荷状態を監視・計測すると共に、隣接サーバ511の負荷状態を問い合わせ、いずれの負荷の方が高いかを判断して、負荷の軽い

サーバを選択する（ステップR2）。ここでは、Webサーバ511の負荷が軽いため、当該Webサーバ511が選択されたものとする。

【0049】次にWebサーバ501は、コンテンツ展開部504により、クライアント端末401からのアクセス要求で指定されたURLに相当するコンテンツデータをデータベース600から検索する（ステップR3）。ここでは、コンテンツ（AAA）601が検索されたものとする。

【0050】この場合、Webサーバ（#1）501は、検索されたコンテンツ（AAA）601を、上記ステップR2で選択されたWebサーバ（#2）511上に、コンテンツ展開部504によりコンテンツ（AAA）512として展開する（ステップR4）。

【0051】次にWebサーバ（#1）501は、検索されたコンテンツ（AAA）601のURLとWebサーバ（#2）511上に展開されたコンテンツ（AAA）512のURLとを結び付け、両URLの対応表を保持する（ステップR5）。

【0052】以降の動作は図2のシーケンスチャートを参照して説明した前記実施形態（におけるステップ6）とほぼ同様である。即ちWebサーバ（#1）501は、この例のように、クライアント端末401から要求されたコンテンツ（AAA）601を隣接するWebサーバ（#2）511上にコンテンツ（AAA）512として展開した場合、URLリダイレクション部503により、クライアント端末401からのアクセス要求中のURLを展開後のコンテンツ（AAA）512を指す新しいURLに変更させ、当該URL変更後のアクセス要求をWebサーバ（#2）511に送出させる。これにより、Webサーバ（#2）511上に展開されたコンテンツ（AAA）512が結果のHTMLとしてWebサーバ（#2）511からクライアント端末401に返信される。なお、クライアント端末401から要求されたコンテンツ（AAA）がアクセス要求先のWebサーバ（#1）501上に展開された場合には、当該コンテンツ（AAA）はWebサーバ（#1）501からクライアント端末401に返信される。

【0053】このように本実施形態においては、負荷の軽いサーバに動的にコンテンツを展開することにより、サーバの負荷を均質化し、システム全体としての応答性能を向上させることができる。また、コンテンツをデータベース600で管理することにより、コンテンツの一元管理が行え、安全性を高めることもできる。

【0054】以上に述べた実施形態では、本発明をWebシステムにおける負荷分散に適用した場合について説明したが、これに限るものではなく、アプリケーションサーバシステムにおける負荷分散にも適用し得る。この場合、負荷の軽いサーバ計算機にアプリケーションプログラムを転送して実行させることにより、システム全体

の性能を向上させることが可能となる。

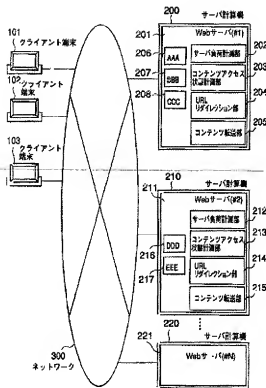
【0055】この他、本発明は、上記各実施形態に限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で種々に変形することが可能である。更に、上記実施形態には種々の段階の発明が含まれており、開示される複数の構成要件における適宜な組み合わせにより種々の発明が抽出され得る。例えば、実施形態に示される全構成要件から幾つかの構成要件が削除されても、発明が解決しようとする課題の全てを述べた課題が解決でき、発明の効果の全てが得られる場合には、この構成要件が削除された構成が発明として抽出され得る。

【0056】

【発明の効果】以上詳述したように本発明によれば、クライアントから要求されたコンテンツの配信を負荷の軽いサーバ計算機で行わせることにより、システム全体のサーバ計算機の負荷を均等化にすることができ、システム全体の応答性能を高めることができる。しかも、負荷分散装置等を使用することなく、サーバ計算機の負荷を均質化できる。

【図面の簡単な説明】

【図1】本発明の第1の実施形態に係るWebシステム



【図1】

のシステム構成を示す図。

【図2】同第1の実施形態における動作を説明するためのシーケンスチャート。

【図3】同第1の実施形態の変形例を説明するためのシーケンスチャート。

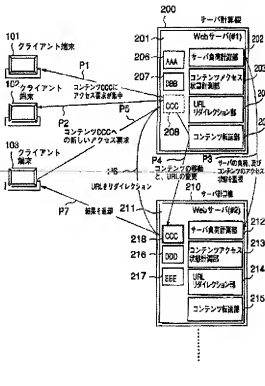
【図4】本発明の第2の実施形態に係るWebシステムのシステム構成を示す図。

【図5】同第2の実施形態における動作を説明するためのシーケンスチャート。

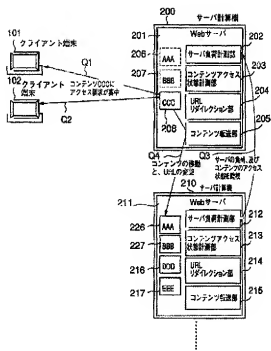
【符号の説明】

101~103, 402, 402...クライアント端末
200, 210, 220, 500, 510, 520...サーバ計算機
201, 211, 221, 601, 511, 521...Webサーバ
202, 212, 502...サーバ負荷制御部
203, 213...コンテンツアクセス状態制御部
204, 214, 503...URLリダイレクション部
205, 215...コンテンツ転送部
504...コンテンツ展開部
600...データベース

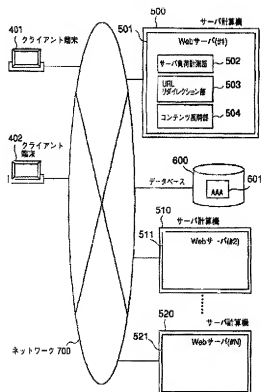
【図2】



【図3】



【図4】



【図5】

